**Course: Advanced Bioinformatics**

**Module title: Progressive alignment methods**

**Module no. : 39**

The approach of progressive alignment is to begin with an alignment of the most alike sequences, and then build upon the alignment using other sequences. Progressive alignments work by first aligning the most alike sequences using dynamic programming, and then progressively adding less related sequences to the initial alignment.

Difficulties with progressive alignments

CLUSTALW and CLUSTALX are progressive alignment programs that follow the following steps:

1. Perform pairwise alignments of all of the sequences
2. Use the alignment scores to produces a phylogenetic tree using neighbor-joining methods
3. Align the sequences sequentially, guided by the phylogenetic relationships indicated by the tree

The initial pairwise alignments are calculated using an enhanced dynamic programming algorithm, and the genetic distances used to create the phylogenetic tree are calculated by dividing the total number of mismatched positions by the total number of matched positions.

Alignments are associated a weight based on their distance from the root node.

In a progressive alignment, gaps are added to a profile of an existing multiple sequence alignment. Statistical tests have been prepared in order to accumulate gaps between secondary structure elements, which models what is found in nature. Such information is incorporated into scoring an alignment using CLUSTALW.

PILEUP is the multiple sequence alignment program that is part of the Genetics Computer Group (GCG) package developed at the University of Wisconsin. In PILEUP, multiple sequence alignment is performed by first aligning each of the sequences in a pair-wise fashion using a Needleman-Wunsch approach. The resulting scores are used to produce a tree by the un-weighed pair-group method using arithmetic averages (UPGMA). The resulting tree is then used to guide the alignment of the most closely related sequences and groups of sequences.

### Problems with progressive alignments

The difficulty with progressive alignments is that they depend upon the initial pair-wise sequence alignments. If the sequences are closely related, then the likelihood is good that the initial alignment contains relatively few errors. However, if the initial sequences are distantly related, then there will be more errors in the alignment, which will propagate through the rest of the alignments.

The second issue is that suitable scoring matrices and gap penalties must be chosen to apply to the sequences as a set.